A New Distribution-Random Limit Normal Distribution

Xiaolin Gong *   Shuzhen Yang †

April 8, 2013

Abstract: This paper introduces a new distribution to improve tail risk modelling. Based on the classical normal distribution, we define a new distribution by a series of heat equations. Then, we use market data to verify our model.

Keywords: fat-tail, peak, normal distribution, heat equation.

1 Introduction

According to Espen Gaarder Haug(2007), Wesley C. Mitchell conducted the first empirical study of fat-tailed (high-peaked) distributions in price data as early as 1915. Following Mandelbrot's famous 1962/63 paper on fat-tails, there have been an abundant research focusing on the non-explained empirical fact. Time-varying volatility, standard returns, mixture model, jump-diffusion model, stochastic volatility, implied distributions are all important tools for risk measurement and management. However, the 2007-08 financial crisis exposed the deficiencies in existing risk models and reinforced the importance of methodology improvements. More detial see [1], [2], [3].

In this paper, we consider a series of heat equations which relate with normal distribution. Assuming the volatility of random varible dependent on the value of random varible, we define a new distribution-random limit normal distribution.

The remainder of the paper followed as: Section 2 introduces a new distribution-random limit normal distribution, and gives two usefull parameter models. Section 3 use the random limit normal distribution to fit history data. Some technique proof is given in the Section 4.

2 A New Distribution

2.1 Random Limit Normal Distribution

Alternatively, we propose a definition of normal distribution:

*Institute for Economics and Institute of Financial Studies, Shandong University, Jinan, Shandong 250100, PR China(agcaelyn@gmail.com).

†School of mathematics, Shandong University, Jinan, Shandong 250100, PR China(yangsz@mail.sdu.edu.cn).

Definition 2.1 *A -dimensional random vector $X = (X_1, \cdots, X_d)$ on a sublinear expectation space $(\Omega, \mathcal{F}, P)$ is called (centralized) normal distribution, if*

$$aX + b\bar{X} \overset{d}{=} \sqrt{a^2+b^2}X, \quad \forall a, b > 0,$$

where $\bar{X}$ is an independent copy of X, which means $\bar{X}$ is independent of X, and $\bar{X} \overset{d}{=} X$. When $d = 1$, we have $X \overset{d}{=} N(0, \sigma^2)$, where $\sigma^2 = E[X^2]$.

Let $u(t,x) = E[\varphi(x + \sqrt{t}X)]$, then $u(t,x)$ satisfies the following heat equation:

$$\partial_t u(t,x) - \frac{1}{2}\sigma^2 \partial^2_{xx} u(t,x) = 0, \quad u(0,x) = \varphi(x), \quad x \in R. \tag{2.1}$$

Set $\varphi(x) = 1_{\{x \le y\}}$, we have

$$E[\varphi(X)] = P(X \le y) = u_y(1,0) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{y} \exp(-\frac{x^2}{2\sigma^2}) dx. \tag{2.2}$$

So the normal distribution of X is

$$F_X(y) = u_y(1,0), \quad y \in R.$$

Following the above definition of normal distribution, we give the definition of the random limit normal function.

Definition 2.2 *For a given X on a probability space $(\Omega, \mathcal{F}, P)$, with function $h(\cdot) : R \longrightarrow (0, +\infty)$, and the heat equation*

$$\partial_t u(t,x) - \frac{1}{2}\sigma^2 h^2(y) \partial^2_{xx} u(t,x) = 0, \quad u(0,y) = 1_{\{x \le y\}}, \quad x \in R.$$

We define the random limit normal function $F_X(\cdot)$ of X as

$$F_X(y) = u_y(1,0) = \frac{1}{\sqrt{2\pi(\sigma h(y))^2}} \int_{-\infty}^{y} \exp(-\frac{x^2}{2(\sigma h(y))^2}) dx, \tag{2.3}$$

where $EX = 0$, $EX^2 = \sigma^2$.

Remark 2.3 *In general, the random limit normal function is not a distribution. However, in the following, we will show two cases when it become a distribution .*

Theorem 2.4 *For a given X on a probability space $(\Omega, \mathcal{F}, P)$, $EX = 0$, $EX^2 = \sigma^2$. For function $h(\cdot) : R \longrightarrow (0, +\infty)$, If any of the following case is right, $F_X(\cdot)$ is a distribution.*

case 1: $h(\cdot)$ is continuous increasing on $(-\infty, 0)$, and continuous decreasing on $[0, +\infty)$;

case 2: Set $h(y) = Ky + c$, $c > 0$, if $y \le 0$, $K < 0$, if $y > 0$, $K > 0$.

Proof. case 1: If $y < 0$, for a given small $\delta > 0$, set $y + \delta < 0$. Then, by the definition of $F_X(\cdot)$, we have

$$\begin{aligned} F_X(y) = & \tfrac{1}{\sqrt{2\pi(\sigma h(y))^2}} \int_{-\infty}^{y} \exp(-\tfrac{x^2}{2(\sigma h(y))^2}) dx, \\ F_X(y+\delta) = & \tfrac{1}{\sqrt{2\pi(\sigma h(y+\delta))^2}} \int_{-\infty}^{y+\delta} \exp(-\tfrac{x^2}{2(\sigma h(y+\delta))^2}) dx. \end{aligned}$$

Note that, $\sigma h(y+\delta) > \sigma h(y)$. By the comparing of two normal distribution, we have

$$F_X(y+\delta) \geq F_X(y).$$

Similar, if $y \geq 0$, we aslo have

$$F_X(y+\delta) \geq F_X(y).$$

So $F_X(\cdot)$ is a increasing function on R, and when $y \longrightarrow +\infty$, $F_X(y) \longrightarrow 1$.

The proof of case 2 is more technique, we will show it in the Appendix. ■

Remark 2.5 *Note that, we don't need the condition $EX = 0$ in the proof of Theorem 2.4. From now, we use the function h in the case 2, i.e. $h(y) = K \cdot y + c$. For reader convenience, we denote $N(u, \sigma^2, K, c)$ as the random limit normal distribution of X.*

3 Two Example

3.1 Test of Hypothesis

Following the classical work of K. Pearson, we conduct goodness of fit test for random limit normal distribution.

When use the method to group the data, the number of groups is m, while n is the sample size. We suppose there are mutually disjoint intervals $I_1, \cdots, I_m$.

$$\lambda = \sum_{i=1}^{m} \frac{n}{p_i}(q_i - p_i)^2 \tag{3.1}$$

where q_i is the frequency of the i th group, $1 \leq i \leq m$, and p_i is the corresponding probability of the fitting distribution.

K. Pearson proves the following theorem:

Theorem 3.1 *If the theoretical distribution is right, when the size of sample $n \longrightarrow \infty$, the limit distribution of statistic λ is the χ^2 distribution in $(\Omega, \mathcal{F}, P)$ with $k-1$ degrees of freedom.*

If the parameters in theoretical distribution are unknown, according to the theorem of Fisher, we take the limit distribution of statistic λ as the χ^2 distribution in $(\Omega, \mathcal{F}, P)$ with $k-1-r$ degrees of freedom. And r is the number of unknown parameters.

3.2 Afitting Example 1

For distributions $N(\mu, \sigma^2)$ and $N(\mu, \sigma^2, K, c)$, we estimate μ and σ by classical method. The fitting result of normal distribution and random limit normal distribution is given. We use the test index δ and δ^{ccn} to compare normal distribution and random limit normal distribution, δ and δ^{ccn} follow as:

$$\delta = \sum_{i=1}^{m} \frac{n}{p_i}(q_i - p_i)^2 \tag{3.2}$$

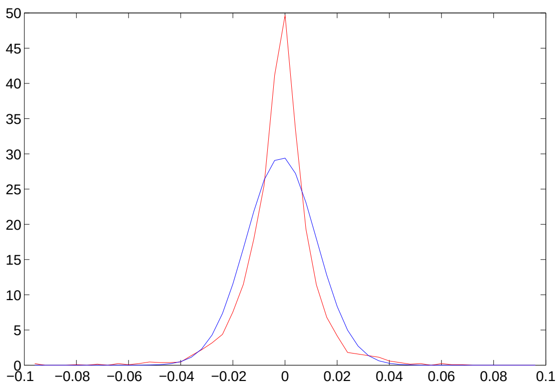
$$\delta^{ccn} = \sum_{i=1}^{m} \frac{n}{p_i} (q_i - p_i^{ccn})^2 \tag{3.3}$$

Where q_i is the frequency of the i th group, $1 \leq i \leq m$, m is the number of group, n is the number of data, p_i is the corresponding probability of the fitting normal distribution, p_i^{ccn} is the corresponding value of the fitting random limit normal distribution.

Using $N(\mu, \sigma^2)$ to fit the data of S@P 500 2000-2012 Log Return, and the parameters is

The style of data	S@P 500
Length of data	3311
μ	$1.70610051464811e-005$
σ	0.0134453931516791
δ	1129.40451908248

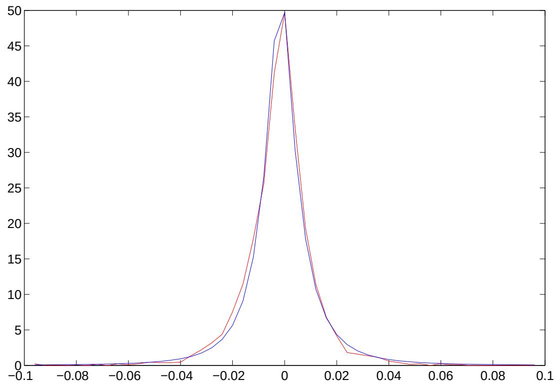
and the fitting figure is



Using $N(\mu, \sigma^2, K, c)$ to fit the data of S@P 500 2000-2012 Log Return, and the parameters is

The style of data	S@P 500
Length of data	3311
μ	$1.70610051464811e-005$
σ	0.0134453931516791
K	$-24(y < 0), 24(y \geq 0)$
c	0.5
δ^{ccn}	349.545332843509

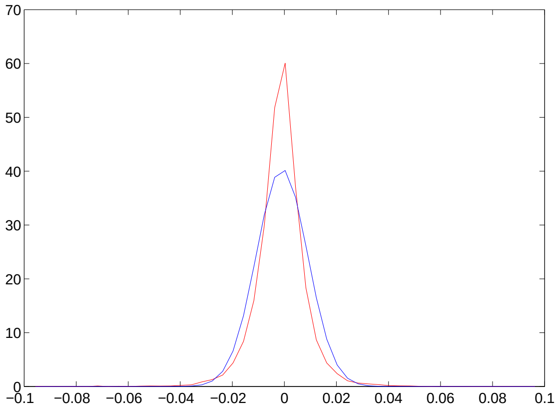
and the fitting figure is



Using $N(\mu, \sigma^2)$ to fit the data of S@P 500 1950-2012 Log Return, the parameters is

The style of data	S@P 500
Length of data	15852
μ	$1.70610051464811e-005$
σ	0.0134453931516791
δ	4830.10362738027

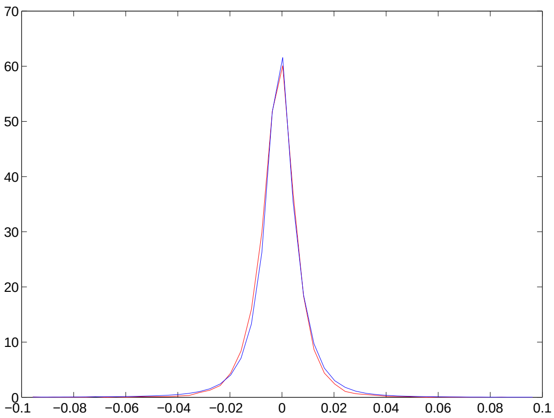
and the fitting figure is



Using $N(\mu, \sigma^2, K, c)$ to fit the data of S@P 500 1950-2012 Log Return, the parameters is

The style of data	S@P 500
Length of data	15852
μ	$1.70610051464811e-005$
σ	0.0134453931516791
K	$-30(y<0), 30(y \geq 0)$
c	0.56
δ^{ccn}	1201.24543468315

and the fitting figure is

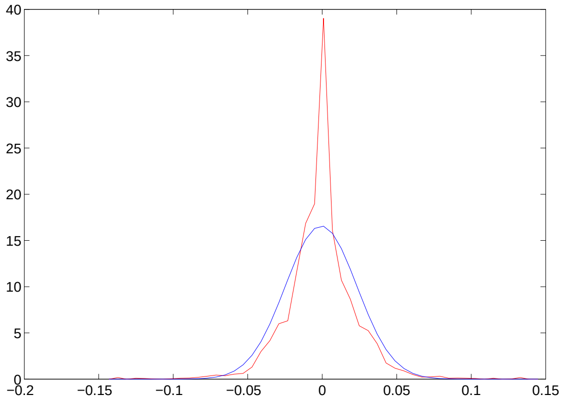


3.3 Afitting Example 2

Using $N(\mu, \sigma^2)$ to fit the data of MSFT 1986-2012 Log Return, the parameters is

The style of data	MSFT
Length of data	15852
μ	$1.70610051464811e-005$
σ	0.0134453931516791
δ	2388.80170177018

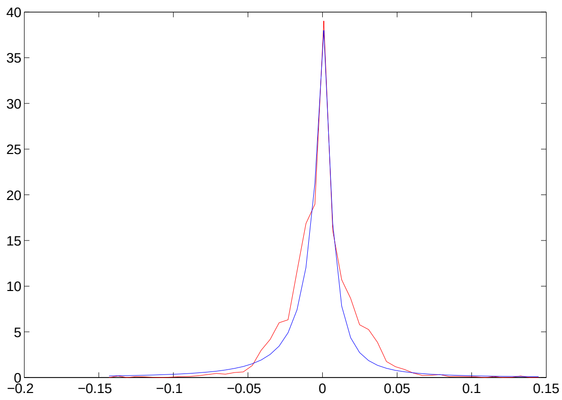
and the fitting figure is



Using $N(\mu, \sigma^2, K, c)$ to fit the data of MSFT 1986-2012 Log Return, the parameters is

The style of data	MSFT
Length of data	15852
μ	$1.70610051464811e-005$
σ	0.0134453931516791
K	$-28.3(y<0), 28.3(y \geq 0)$
c	0.24
δ^{ccn}	1201.24543468315

and the fitting figure is



Recent research is strongly motivated by the financial meltdown to develop models and methods to better predict extreme events. In this paper, we propose a new distribution: continuous change normal distribution which includes the normal distribution as a special case

and takes into account of main statistically relevant stylized facts of asset returns data: fat-tail, high-peak.

4 Appendix

The proof of case 2 in Theorem 2.4:

Proof. case 2: For convenience, we set $\sigma = 1$.

when $y \in (-\infty, 0]$, $K < 0$. In order to prove $F_X(\cdot)$ is a distribution, we need to show that $F_X(\cdot)$ is a continuous and increasing to $1'$s function. By the definition of $F_X(\cdot)$, we have

$$F_X(y) = \frac{1}{\sqrt{2\pi(Ky+c)^2}} \int_{-\infty}^{y} \exp(-\frac{x^2}{2(Ky+c)^2}) dx.$$

Considering the right derivative function of $F_X(\cdot)$, and denote as $f_X(\cdot)$.

$$\begin{aligned} f_X(y) = & \frac{K}{\sqrt{2\pi(Ky+c)^4}} \int_{-\infty}^{y} [\exp(-\frac{x^2}{2(Ky+c)^2}) \cdot (\frac{x^2}{(Ky+c)^2} - 1)] dx \\ & + \frac{1}{\sqrt{2\pi(Ky+c)^2}} \exp(-\frac{y^2}{2(Ky+c)^2}). \end{aligned} \tag{4.1}$$

Then, we have

$$\begin{aligned} f_X(0) = & \frac{1}{\sqrt{2\pi(Ky+c)^2}} - \frac{K}{2(Ky+c)} + \frac{K}{2(Ky+c)} \\ = & \frac{1}{\sqrt{2\pi(Ky+c)^2}} \\ > & \ 0. \end{aligned}$$

Set

$$\begin{aligned} g_X(y) = & \int_{-\infty}^{y} \frac{K}{Ky+c} [\exp(-\frac{x^2}{2(Ky+c)^2}) \cdot (\frac{x^2}{(Ky+c)^2} - 1)] dx \\ & + \exp(-\frac{y^2}{2(Ky+c)^2}). \end{aligned}$$

Set $m(y) = \frac{y}{Ky+c}$, then $m \in (\frac{1}{K}, 0]$. By parameter transformation, we have

$$\tilde{g}_X(m) = \exp(-\frac{m^2}{2}) + \int_{-\infty}^{m} K[\exp(-\frac{z^2}{2}) \cdot (z^2 - 1)] dx.$$

Note that, m is a increasing to $0'$s function of y. Also we have $\tilde{g}_X(0) > 0$. The derivative function of $\tilde{g}_X(\cdot)$ is

$$\begin{aligned} \tilde{g}'_X(m) = & -\exp(-\frac{m^2}{2}) \cdot m + K[\exp(-\frac{m^2}{2}) \cdot (m^2 - 1)] \\ = & \exp(-\frac{m^2}{2}) \cdot (Km^2 - m - K). \end{aligned}$$

For $m \in (\frac{1}{K}, 0]$, so $\tilde{g}'_X(\cdot) > 0$. Then $\tilde{g}_X(\cdot)$ is increasing function of m, and also $g_X(\cdot)$ and $f_X(\cdot)$ is a increasing function of y.

By the equation (4.1), $f_X(y) \longrightarrow 0$, when $y \longrightarrow -\infty$. So $f_X(y) > 0$, $y \in (-\infty, 0]$, and $F_X(\cdot)$ is a increasing function on $(-\infty, 0]$.

Similar, we could prove $F_X(\cdot)$ is a increasing function on R, and when $y \longrightarrow +\infty$, $F_X(y) \longrightarrow 1$.

This complete the proof. ■

5 References

References

[1] Espen Gaarder, Haug., (2007), Derivatives Models on Models, The Wiley Finance Series.

[2] Inui, K. and Kijima, M., (2005), "On the significance of expected shortfall as a coherent risk measure", Journal of Banking and Finance, 29, 853-864.

[3] Knight, Frank., (2006), Risk, Uncertainty, and Profit, Mineola: Dover Publications(republication of the 1957 edition of the work originally published in 1921 by the Houghton Mifflin Company).

[4] Mises, Ludwig von., (1966), Human Action: A Treatise on Economics, Chicago: Regnery.

[5] Mandelbrot, B., (1963),The variation of certain speculative prices, Journal of Business 26, 394-419. Mandelbrot, B., (1963), The variation of certain speculative prices, Journalof Business 26, 394-419.